

DOCUMENT RESUME

ED 206 712

TM 810 615

AUTHOR Marston, Paul T.; Zimmerer, Linda K.
TITLE The Effect of Using More than One Type of Teacher Observation System in the Same Study.
INSTITUTION Texas Univ., Austin. Research and Development Center for Teacher Education.
SPONS AGENCY National Inst. of Education (DHEW), Washington, D.C.
PUB DATE Sep 79
NOTE 31p.
EDRS PRICE MF01/PC02 Plus Postage.
DESCRIPTORS Analysis of Variance; *Bias; *Classroom Observation Techniques; Classroom Research; *Data Collection; Elementary Secondary Education; Teacher Behavior; *Test Reliability; Videotape Recordings
IDENTIFIERS Classroom Observation Record; Classroom Observation Scales; *Coding; Flanders System of Interaction Analysis

ABSTRACT

The multiple use of several classroom observation instruments was investigated to determine its influence upon the way an observer records behavior, since such an influence may cause discrepancies in findings between studies. Twenty-four volunteer students were divided into three groups and trained to use one or more of three classroom observation instruments: the first group was trained to use the low-inference (requires little observer judgment) Flanders System of Interaction Analysis; group 2 was trained to use the high-inference (requiring high observer judgment) Classroom Observation Record and Classroom Observation Scales; group 3 was trained in all three. Observers then matched and coded videotapes of classroom action in a fifth grade social studies session. Experiments were then conducted to test the influences of high and low-inference coding techniques upon one another. Analysis showed an influence both ways, with coders bringing their high and low-inference codings into consistency with one another; the effect of the low-inference on the high-inference being strongest. Examination of the specific items changed indicated that overlap in content between the systems probably accounts for most changes. The results pointed to a possible source of bias in process-product studies that use multiple observation systems to maximize data collection. (Author/AEP)

* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *

☒ This document has been reproduced as
received from the person or organization
originating it.

☐ Minor changes have been made to improve
reproduction quality.

• Points of view or opinions stated in this docu-
ment do not necessarily represent official NIE
position or policy.

"PERMISSION TO REPRODUCE THIS
MATERIAL HAS BEEN GRANTED BY

C. Walton

TO THE EDUCATIONAL RESOURCES
INFORMATION CENTER (ERIC)."

The Effect of Using More Than One Type of Teacher Observation System in the Same Study

Paul T. Marston and Linda K. Zimmerer

Many of the current prescriptions for effective teaching are based on studies in which classroom behavior was observed, quantified, and related to student performance. In such studies, one or more observers watch a teacher's class and record what takes place using a coding instrument. A recent review of a large number of such studies has shown that a number of teacher behaviors recorded by these instruments were related to student achievement gain (Medley, 1977). The patterns of these relationships were not always simple, often being complicated by other factors, such as socioeconomic status, grade level, or subject matter. Because of this complexity and because most systems were designed only to record information relevant to specific theories of teaching, no single system is likely to record many of the behaviors that might be associated with student success. Thus, a common tendency in research has been to use several observation systems within a single study to encompass a larger range of relevant behaviors. This type of approach has been suggested by Stallings in the discussion of the Follow-Through evaluation, (1977). The report emphasized that the variety of goals and methods used in the different programs being evaluated required the collection of many kinds of data. It stated that there was a need for instruments that describe the totality of events in the classroom--the environment, the materials, the activities of the teachers and children, and the

ED206712

TM810615

Abstract

The biasing effects of an observer using both high- and low-inference coding systems were tested. Volunteer coders were trained to use either one or both types of systems and then coded a set of video tapes. The analysis showed an influence both ways, with the effect on the high-inference by the low-inference being strongest. Examination of the specific items changed indicated the overlap in content between the systems probably accounts for most of the changes. The results pointed to a possible source of bias in process-product studies that use multiple observation systems to maximize data collection.

interactions that occur between them.

The advantage of collecting as many kinds of observational data as possible in a single study is clear. The availability of classrooms to observe is limited and the cost of observers to send is high, so getting the maximum from each of them is imperative. But what also needs to be considered is whether there are any detrimental effects. One problem is that a large number of observational variables makes drawing inferences more difficult. This problem can be dealt with in several ways, such as increasing the number of independent observations or by demonstrating the findings in more than one study. A potentially more serious problem concerns whether the use of several observation systems changes the way an observer records things compared with using only one system at a time. If this occurs, it may cause systematic but unknown changes in the findings from study to study that would depend upon the particular cluster of instruments used in a particular study.

The potential for this type of bias might depend, in part, on the similarity of the two instruments being used. If they are very similar, then very little bias would be expected since essentially the same information is being recorded in two different places. Conversely, if the two instruments were very different in the information they recorded, then bias might not be expected, either, because the observation acts would be independent. The most likely situation for bias would be those cases in which the two instruments recorded the same behavior in different ways. An example would be recording teacher talk, in

which one instrument required timing the amount of talk, while the other one required it to be rated on a scale from "very little" to "very much."

Conceptually, observation instruments can be classified as using either sign or category recording and as requiring either high- or low-inference judgments. In a sign instrument, only the presence or absence of a behavior is recorded, while in a category system the number of times a behavior occurs is recorded. A sign system may be modified into a rating scale, which is often used to estimate the frequency of occurrence of a behavior. The classification as either high- or low-inference depends on the amount of judgment required by the coders. Generally, low-inference systems are also categorical ones, while high-inference ones are usually modified sign systems (Borich, 1977). In the studies that have used multiple observation techniques there are almost always both high- and low-inference instruments used. For example, in a year-long study of junior high math and English classes, a combination of high- and low-inference coding systems were used to predict students' achievement gains (Brophy & Evertson, 1978; Anderson, Evertson & Brophy, 1978). In another study involving concept instruction, high- and low-inference systems were used to measure approach, affect, and on-task behavior (Armento, 1977). In each instance, the low-inference systems were used to record overt teacher behavior, while the high-inference systems were used to record global characteristics such as teacher warmth.

To test the hypothesis that different type systems would

affect one another, observers in the following study were trained to use either high-inference, low-inference, or both types of coding systems. The single low-inference system used was a category system, while the two high-inference systems used were modified sign systems. It was hypothesized that the codings for a given method would be different in the group using both methods compared with that method used alone.

Method

Subjects

The coders in this study were twenty-four student volunteers who responded to an advertisement in the campus newspaper. They were paid \$10 plus \$3 per hour for training time and subsequent experimental time to use one or more of three classroom observation instruments--one low-inference and two high-inference systems. The total payments ranged from \$50 to \$70, depending on the amount of training required. While some of the individuals were either education or psychology majors, none of them had ever coded classroom behavior before.

Training

The coders were randomly assigned to one of three groups of eight persons when they applied for the study. The observers in Group 1 were given seven hours of training on the Flanders System of Interaction Analysis (FSIA) (Flanders, 1970), a low-inference, category system. The coders in Group 2 were given three hours of training on the Classroom Observation Record (COR) (Ryans, 1960), and the Classroom Observation Scales (COS) (Emmer, 1971), two high-inference rating scale systems. The coders in Group 3 were

given training in both types of systems, for a total of ten hours. The training occurred on three consecutive evenings, with Groups 1 and 3 being trained on the FSIA system during the first two evenings, and Groups 2 and 3 being trained on the COS and COR on the third evening. As a consequence of sharing their training with the other two groups, Group 3 members received training on the low-inference system before the high-inference systems. The training was conducted by a coder who was experienced in that system, using training methods provided by the system authors. In some cases minor modifications were made in the training because of time constraints. For example, in the COR and COS training, coders were urged to develop their own set of ground rules for categories through a method of group consensus, rather than using those established by an "expert" coder as suggested by the authors. Of the original 24 coders, two in Group 3 failed to complete the training, leaving that group with only six members.

Design

Each coder observed the videotapes in the same order, with a one-hour period allotted for each of the eight teachers. During the hour, three 10-minute segments of the teacher's class were shown, separated by a 5- to 10-minute break. The coders using the FSIA made their tallies while the tape was being shown, while the coders using COS and COR made their ratings during the break after each segment. The observers worked in pairs, which were assigned based on the individuals having free time in common. The videotape segments were taken from a set of fifth-grade social studies lessons made for another study (Haak et al.,

1971).

To estimate the reliability of the observers for each system, the reliability for every possible coder pair was calculated, transformed, and averaged. These average reliabilities on the first segment of the first tape were .61 for FSIA (ARCSIN transformation, proportion agreement, Scott, 1955), .27 for COR (Fisher's r to z transformation, intraclass correlation with anchor points, Winer, 1971, pp. 289-293), and .66 for COS (Fisher's r to z transformation, intraclass correlation with anchor points).

Analyses

For purposes of analysis this study can be considered as two separate experiments, each involving a control versus a treatment group. The treatment group in both cases consists of coders in the group who used both high- and low-inference observation systems--Group 3, while the control group consists of coders in the groups who used only one type of system. To test whether a low-inference system causes changes in coding a high-inference one, the appropriate control group is the one where the coders used only high-inference coding--Group 2. Conversely, to test the influence of high-inference coding on low-inference coding the control would be Group 1, the low-inference coders.

Scoring

The FSIA observers coded the activity in the classroom into one of ten mutually exclusive categories at a rate of approximately once every three seconds, which produced about 200 codes in each ten-minute segment. To avoid spurious effects

which might be caused by differences in coding rates, the number of codes in each category was converted into the proportion of codes in the segment. In addition to the proportions, some derived measures provided by the author (Flanders, 1970) were added as additional dependent variables. These measures, with the combination of categories making them up shown in parentheses were:

- A. I/d ratio $(1+2+3)/(6+7)$,
- B. I/D ratio $(1+2+3+4)/(5+6+7)$,
- C. Proportion teacher response $(1+2+3)/(1+2+3+6+7)$,
- D. Proportion teacher question $(4)/(4+5)$,
- E. Proportion pupil initiation $(9)/(8+9)$, and
- F. Proportion teacher talk $(1+2+3+4+5+6+7)/(\text{total})$.

The proportion of codes in each category and the six derived measures were calculated for each coder during each segment and used as dependent variables.

Both the COS and COR are rating scale instruments which were coded after each segment. The COS consists of twelve items which are rated from one, meaning "little" of the behavior was observed to five, meaning "much" of the behavior was observed. No factor analysis was available for COS, but it was possible to divide the items into two unit-weighted scales-- those items which were judged to be logically related to the FSIA and those items that were not. These scales, with the items making them up shown in parentheses were:

- 1) FSIA related $(4+7+9)$ and
- 2) Not FSIA related $(1+2+3+5+6+8+10+11+12)$.

Since these two scales were derived after the data had been collected, any findings involving them should be treated as tentative and subject to further experimental verification.

The COR uses a 22-item polar adjective scale which runs from one, for the teacher being closest to the first word in the pair, to seven, for being close to the second word. Four factors for the COR have been identified by its author (Ryans, 1960). Each factor was calculated by forming a unit-weighted sum of the items showing the highest factor loadings. These factors, with the items making them up shown in parentheses were:

- A. Kindly-Understanding (6+7+8+9+20),
- B. Systematic-Organized (14+15+16+18),
- C. Stimulating-Imaginative (10+11), and
- D. Pupil Related Behavior (1+2+3+4).

Notice that seven of the items were not included in any factor (i.e., 5, 12, 13, 17, 19, 21, and 22).

Results

The dependent measures, whether categories, items, or scales, were analyzed separately using unweighted means univariate analyses of variance. Each one used a mixed model design with between-subjects factor treatment group (2) and within-subjects factors teacher (8) and occasion (3).

Only two FSIA categories and one derived measure showed any significant changes as a result of being used with the high-inference systems. The results of this ANOV are shown in Table 1, which, for simplicity, lists only those lines of the source table which relate to the experimental treatment, factor

G. The category "Accepts Feelings" showed a significant treatment by teacher interaction [$F(7,35)=2.74$; $p<.05$], while the category "Asks Questions" showed a significant treatment by teacher by segment interaction [$F(7,35)=2.22$, $p<.05$]. The latter category also showed an almost significant treatment by teacher interaction [$F(7,35)=2.22$, $p<.06$]. The mean proportions of codings for the two interactions are shown in Table 2. As indicated in the column labeled percentage, the "Accepts Feelings" category is seldom coded, accounting for only 0.02% of all codes. In fact, six out of eight teachers were never observed to use this behavior. The interaction resulted from the high plus low group using it for one teacher while the other group did not. The mean proportions for the "Asks Questions" category were slightly, but not significantly, higher for the group using both types of systems, with half the teachers receiving more codings from this group. "Asks Questions" is a common code, with about 18% of the codings occurring in this category. The pattern of means for the treatment by teacher by segment interaction in this category did not lend itself to any simple interpretation. The "Teacher Questioning Ratio" [$4/(4+5)$] showed a significant treatment by tape segment interaction [$F(2,10)=4.03$; $p<.05$]. This was due to a higher ratio of questions for coders in the treatment group during the second ten minute segment. The "Indirect-Direct Ratio," I/D [$(1+2+3+4)/(5+6+7)$], showed a significant three-way interaction of treatment by teacher by segment [$F(14,70)=1.85$; $p<.05$], but again, the mean relationships had no obvious pattern.

The number of FSIA tallies in each category for group 3 were correlated with the observers' ratings on the COS and COR for all 24 tape segments. These correlations are shown in Tables 7 and 8. Because of the repeated measures design, a standard significance test could not be done on these correlations, but it is interesting to note the pattern of "important" correlations--defined as those correlations equal to or exceeding .10 (i.e., 1 % of the variance in common for the two items). Of the 120 possible correlations with COS, 55% met the 1% criterion, with 35 being negative and 31 positive. Out of 220 possible correlations with COR, 55% met the 1% criterion, with 56 being negative and 64 being positive. When only the FSIA items showing significant effects in the ANOV are considered, COR showed one positive relationship with FSIA category 1, while COS showed two negative and one positive one. In both cases, it was the FSIA category with the fewest important relationships. FSIA category 4 showed six negative and three positive correlations with COS items out of a possible twelve. COR showed 15 negative and 2 positive relationships out of the possible 22.

In the COS analysis, significant main effects were found for two items and for one factor. The item "Percent Teacher Presentation" had a mean rating of 1.67 in the control group and 2.40 in the treatment group, which was highly significant [$F(1,5)=27.11$; $p<.01$]. The item "Negative Affect" showed a smaller effect in the opposite direction [$F(1,5)=7.69$; $p<.05$], with mean ratings of 1.35 and 1.13 for the control and treatment groups, respectively. Factor B, based on the sum of those items

logically related to FSIA, also showed a main effect [$F(1,5)=6.62$; $p<.05$], with the control group which used only high-inference instruments scoring an average of 6.36 compared to 7.48 for coders using both systems. The two significant interactions for COS involved treatment by tape segment. One was for "Percent Teacher Presentation" [$F(2,10)=9.86$; $p<.01$], which showed a tendency for the difference between the two groups to get smaller for the second and third segments in each session. The item "Positive Affect" also showed a treatment by segment interaction [$F(2,10)=5.08$; $p<.05$], but the change in ratings came between segments one and two. The COS item with a significant main effect, "Percent Teacher Presentation," showed important correlations with all of the FSIA categories except "Criticizing or Justifying Authority." As shown in Table 7, all but one correlation was negative, the sole exception being with "Lecturing." The other main effect, "Negative Affect," showed only four important correlations, with three of these being positive and one negative. "Positive Affect," which had a significant two-way interaction, showed seven important correlations--four negative and three positive.

Significant main effects were found in the COR analysis for the item "Stereotyped-Original" [$F(1,5)=7.10$, $p<.05$] and factor C, "Stimulating-Imaginative" [$F(1,5)=6.39$, $p<.05$]. In both cases the mean ratings were higher for the group also making low-inference codings--from 4.57 to 5.33 on "Stereotyped-original" and from 9.38 to 10.81 for factor C. This item had important correlations with six FSIA categories, with four being

positive, and two negative, as shown in Table 8. Six items showed treatment by teacher interactions that were significant, and these appeared to fall into two types of patterns as shown in Table 6. In the first pattern most of the teachers were rated slightly higher when both systems were used (at most .57 points), while two or three teachers would show an opposite pattern of being rated lower (up to .86 points). The items showing this pattern were "Confident" (student) [$F(7,35)=2.94$; $p<.05$]; "Erratic-Steady" [$F(7,35)=3.33$; $p<.01$]; "Excitable-poised" [$F(7,35)=2.66$; $p<.05$]; and factor B, "Systematic-Organized" [$F(7,35)=2.96$; $p<.01$]. The other interaction showed a reverse pattern, with five of the eight teachers receiving higher ratings when only high-inference instruments were used. These items were "Uncertain-Confident" (teacher) [$F(7,35)=2.51$; $p<.05$], and "Disorganized-Systematic" [$F(7,35)=2.56$; $p<.05$]. All of these ratings except "Uncertain-Confident" (student) showed at least four important correlations (see Table 8). There were also two treatment by teacher by segment interactions which involved the student scales "Apathetic-Alert" [$F(14,70)=2.14$; $p<.05$] and "Obstructive-Responsible" [$F(14,70)=1.91$; $p<.05$]. The correlations between the two high-inference systems were fairly high in Group 3, with 67% of the 264 possible correlations showing important relationships. These correlations are shown in Table 9.

Discussion

When this experiment was designed, it is not clear what type of bias would be produced by one coding instrument or

another. All that was known was that anecdotal reports from coders indicated their perceptions of teachers might have been affected by using a specific type of coding system. In particular, these coders noted that their training caused them to pay attention to aspects of the classroom that they might have otherwise missed. Only a weak test of the hypothesis could be made--that some sort of influence would be exerted on one coding system by another. Furthermore, it was not known what items or categories would be changed. Thus, when considering the results, it is more important that some changes occurred than what specific items were affected or what form the change took. With the possible exception of COR, the observers were well trained in the systems and should have recorded things in much the same way. Any departure from consistent coding on any item must be considered as a fault in the coding system. The results provide support for the hypothesis that changes in codings would occur when different types of observation systems were used by the same observers. It was possible for either the high-inference systems to affect the low-inference one, or vice versa, and, in fact, changes in coding were observed on both types of systems. The changes were larger on the high-inference systems, and the pattern of changes was also much more systematic for these systems. The proportion of items or categories that were significantly affected was not large for any system. Moreover, the pattern of change for some items was subtle, with the size and direction of the effect depending on which particular teacher was observed.

For FSIA, the group of coders who used both types of observation instruments recorded significantly different coding patterns for some teachers on two of the categories: "Accepts Feelings" and "Asks Questions." The former category is rarely coded, as indicated by only two of the eight teachers ever receiving even a single tally for it. Because the high-inference systems force the observers to attend to affect, it is reasonable that the FSIA coders who also used high-inference coding instruments were more aware of the instances where teachers showed consideration toward students and more ready to code it. Even though the incidence of this behavior is quite low, a humanistic approach to teaching would suggest it may be very important when it does occur. The group using both types of systems also coded four of the teachers as asking considerably more questions than did the group using the low-inference system only. The pattern of correlations with COS indicates coders may have seen these teachers as exercising more control over the class and allowing less pupil-initiated behavior. This was shown by negative correlations with "Pupil Attention," "Percent Teacher Presentation," "Positive Affect," "Task Orientation," "Clarity," and "Enthusiasm," and by positive correlations with "Teacher Initiated Problem Solving," "Negative Affect," and "Passive Pupil Behavior." The questioning category showed thirteen negative correlations with the COR teacher scales, and only one positive one. It appears that those teachers who are perceived as exerting control over the class are seen as doing this partially through a question and answer mode of instruction.

On the whole, effect of the high-inference systems on the low one was not great. The effects that did appear were in the form of interactions, indicating that the direction of any influence on coding depends on the characteristics of the teacher being observed. A possible explanation for the complex findings with FSIA is that there is no clearly defined pattern of codings that describes a good teacher. Therefore, while observers indeed may have been biased, the direction of this bias was not predictable. Another explanation is that because the coding of a teacher on the low-inference system always preceded the high-inference coding, the direction of influence also would be expected to go from low- to high-inference.

The COS system showed the highest percentage of direct changes (17%) as a result of being used in conjunction with a low-inference system. Overall changes on ratings between the groups were found for the categories "Percent Teacher Presentation" and "Negative Affect." The coders who used both types of coding system saw more teacher presentation and less negative affect. These coders probably saw more teacher presentation since they were required to attend to the FSIA category of "Teacher Lectures" and there was, in fact, a large correlation ($r = .58$) between the ratings and the category. The "Negative Affect" scale correlated positively with the FSIA scale "Criticizes or Justifies Authority" and negatively with "Praises or Encourages," as would be expected. As with lecturing, it seems reasonable that the low-inference coding made the observers more aware of how often such behaviors occurred. These observers

saw less positive affect in the teacher's first tape segment, but not in the others. Likewise, the difference in rated teacher presentation between the two observer conditions was less for the later segments. Two composite scores were analyzed for the COS items, the first of which consists of three items which were judged to overlap FSIA. The coders who used both types of systems showed significantly higher scores on the overlapping composite. It appears that coders who used both COS and FSIA brought their judgments on the COS into line with what they recorded while watching the teachers. Things such as "Criticizing" probably stand out in ordinary classroom procedure, but when it has to be recorded as it occurs, the low rate of incidence becomes apparent. Conversely, the large amount of time the teacher spends lecturing also becomes apparent when it is recorded as it happens.

Of the three, the COR was affected the most when it was used in conjunction with a low-inference system. These results must be treated with skepticism, however, because the interrater reliability was so poor for the COR. The coders, using both types of systems rated teachers as being more alert and higher on the factor "Stimulating-Imaginative," which included the alertness scale. Alertness was positively correlated with FSIA categories "Praises or Encourages", "Giving Directions," and "Student-Talk-Initiation." It was negatively correlated with "Asks Questions," "Lecturing," and "Student Talk Response." A possible explanation is that the attention to the details of the teacher's behavior required by FSIA makes it clear how many

different things a teacher is doing, no matter how dull he or she may appear to be. Another pattern that occurred for four of the items and for one factor was that observers rated from three to five of the eight teachers much lower when also using the FSIA. They rated the other three teachers higher, but not quite as much. These categories included "Erratic-Steady," "Excitable-Poised," "Uncertain-Confident," and "Disorganized-Systematic." The factor "Systematic-Organized", which includes all of the previously listed items except "Uncertain-Confident", showed the same pattern. These items generally showed positive correlations with the FSIA categories "Praises or Encourages," "Accepts or Uses Student Ideas," and "Silence or Confusion." They showed negative correlations with "Asks Questions," "Criticizing or Justifying Authority," and, to a lesser extent, "Lecturing." Three of the teachers were coded lower by the group using both systems, which indicated the same influence may be working for each item. Since all of these items involve fairly overt behaviors, it is possible the combined coders were more aware of the lack of occurrence of affective behavior in the teachers. A related pattern occurred for the student behavior "Uncertain-Confident," which was positively related to "Questioning" and negatively related to "Praises or Encourages," and to "Criticizing or Justifying Authority." An attention explanation is probably appropriate here, too. As with the COS, coders were much more aware of how often behaviors occurred since they were actively tallying these as compared to the coders using high-inference only, who were just watching the

videotapes. The COR is a more affectively oriented system than the COS, and it showed more correlations with the FSIA categories with affective components. It is interesting to note that, as the experiment progressed, these coders gave progressively lower ratings to each teacher on the "Systematic- Organized" factor. Perhaps this verifies the old adage "familiarity breeds contempt."

Conclusions

The hypothesis that the act of coding one type of observation system can influence the coding of another system was partially supported. Most of the influence was on the more subjective high-inference systems. This was to be expected, since more judgment is required to make the high-inference codings, and, therefore, more opportunity exists for influence to take place. There were some changes in the low-inference codings, and there may have been others that were masked because the influence varied from observer to observer. The overall pattern was one of the observers bringing their high- and low-inference codings into consistency with one another. This was most clearly evident by the increase in the high-inference ratings for teacher talk and the decrease in the high-inference ratings for affective behaviors when used with the low systems. The low-inference category, "Questioning," which was associated with negative high-inference items, was influenced the most by the high systems. It is clear that one cannot indiscriminately mix methods of observation and expect them to work as if each were being used alone. The economies of using the same observers

for several data collection tasks need to be weighed against the probable distortion of the data collected.

The more far-reaching interpretation of these results relates to the issue of the reliability of classroom observation as the primary measuring instrument in teacher improvement research. The influence of other data collection procedures is only a small part of the many possible biasing sources that could distort classroom observations. The impact of these biasing factors is, for the most part, unknown at this time. If they are large, then they could seriously distort the data in a process-product study which, in turn, could lead to erroneous conclusions about the relationships between teacher behavior and student outcomes. This study certainly lends credibility to that possibility.

References

- Anderson, L.M., Evertson, C.M., & Brophy, J.E. Analyses of treatment effects in an experimental study of first grade reading groups. Paper presented at Annual Meeting of AERA, Toronto, Canada, March 1978.
- Armento, B.J. Selected correlates of effective teacher behavior during concept instruction: Their design, utility and limitations. Paper presented at Annual Meeting of AERA, New York City, April, 1977.
- Borich, G.D. The appraisal of teaching. Concepts and process. Reading, Massachusetts: Addison-Wesley Publishing Co., 1977.
- Brophy, J.E., & Evertson, C.M. Process-outcome relationships in the Texas Junior High School Study: Low-inference coding of specific behaviors. Paper presented at the Annual Meeting of AERA, Toronto, Canada, March, 1978.
- Emmer, E.T. Classroom Observation Scales. Austin, Texas: The Research and Development Center for Teacher Education, 1971.
- Flanders, N.A. Analyzing teaching behavior. Reading, Mass.: Addison-Wesley Publishing Co., 1970.
- Haak, R.A., Lawrence, W.A., & Peck, R.F. ITEC: An experiment in school consultation. Publication No. 1082, The Research and Development Center for Teacher Education, The University of Texas at Austin, 1971.
- Medley, D.M. Teacher competence and teacher effectiveness: A review of process-product research. Washington, D.C.: American Association of Colleges for Teacher Education, August, 1977.
- Ryans, D.G. Characteristics of teachers. Washington, D.C.: American Council on Education, 1960.
- Scott, W.A. Reliability of content analysis: The case of nominal scale coding. The Public Opinion Quarterly, 1955, 19, 321-325.
- Stallings, J.A. The importance of multiple data collection instruments when describing the educational process. Paper presented at the Annual Meeting of AERA, New York City, April 1977.
- Winer, B.J. Statistical principles in experimental design (2nd ed.). New York: McGraw-Hill, 1971.

TABLE 1
ANALYSES OF VARIANCE ON PROPORTIONS OF CODES IN EACH
CATEGORY FOR THE FLANDERS SYSTEM OF INTERACTION ANALYSIS (FSIA)

Scale	% (1)	Effect (2) (degrees of freedom)			
		G (1,5)	GT (7,35)	GS (2,10)	GTS (14,70)
1 Accepts Feelings	.02	4.71	2.74*	.14	.42
2 Praises or Encourages	1.27	.47	.79	1.38	1.32
3 Accepts or Uses Ideas	10.23	.78	.26	2.09	.51
4 Asks Questions	18.13	.11	2.22	2.68	2.16*
5 Lecturing	24.50	.13	.32	1.22	.89
6 Giving Direction	11.13	.24	.89	1.79	1.62
7 Criticizing or Justifying Authority	.11	.01	.21	1.44	.33
8 Student Talk-Response	24.26	5.42	1.20	2.27	.52
9 Student Talk-Initiation	4.54	1.98	.64	1.72	1.23
10 Silence or Confusion	5.82	2.33	.54	.02	.81
Derived Scores (3)					
A I/d Ratio		2.16	1.33	1.08	1.37
B I/D Ratio		.04	1.11	3.43	1.85*
C Teacher Response Ratio		.71	.68	.80	.74
D Teacher Question Ratio		.14	.85	4.03*	.92
E Pupil Initiation Ratio		2.22	.61	1.70	1.20
F Teacher Talk Proportion		.01	.51	2.65	.70

* $p < .05$ ** $p < .01$

Note:

(1) % = The percentage of all FSIA responses that fell in the particular category.

(2) G = Group: (1) Low-Inference or (2) Low- + High-Inference

T = Teacher

S = Segment

(3) See text for explanation of derived scores.

TABLE 2
 MEANS FOR SIGNIFICANT INTERACTIONS ON THE LOW-INFERENCE
 FLANDERS SYSTEM OF INTERACTION ANALYSIS (FSIA)

Condition	1	2	3	Teacher 4	5	6	7	8
(1) Accepts Feeling								
Low	0	0	0	0	0	.021	0	0
Low(+High)	.028	0	0	0	0	.21	0	0
(4) Asks Questions								
Low	21.7	15.7	17.7	18.9	14.4	22.0	19.1	12.7
Low(+High)	20.8	18.3	16.8	18.8	13.7	22.6	22.0	14.7

Condition	1	Tape Segment 2	3
(D) Questioning Ratio			
Low	42.6	42.6	49.3
Low(+High)	41.8	47.8	48.9

TABLE 3
Analyses of Variance of Ratings for
the Classroom Observation Scale (COS)

Scale	Effect (1) (degrees of freedom)			
	G (1,5)	GT (7,35)	GS (2,10)	GTS (14,70)
1 Pupil Attention	.77	.52	.58	1.78
2 Teacher-Initiated Problem Solving	.14	.54	.13	.38
3 Pupil to Pupil Interaction	4.06	1.23	.71	1.04
4 Percent Teacher Presentation	27.11**	.82	9.86**	1.06
5 Negative Affect	7.69*	2.12	1.59	1.02
6 Positive Affect	.12	1.04	5.08*	1.41
7 Higher Level Cog- nitive Pupil Behavior	1.19	.80	2.15	.83
8 Passive Pupil Behavior	2.61	.24	3.15	1.62
9 Convergent Evaluative Behavior	.02	.16	.63	1.17
10 Task Orientation	.69	1.54	1.72	.97
11 Clarity	.02	.99	.01	.53
12 Enthusiasm	1.98	1.32	.09	1.39
Factors (2)				
A FSIA Related	6.62*	.53*	.05	1.60
B Not FSIA Related	.19	.19	.99	.71

* $p < .05$ ** $p < .01$

Note:

(1) G = Group: (1) Low-Inference or (2) Low- + High-Inference

T = Teacher

S = Segment

(2) See text for explanation of factor scores.

TABLE 4
 MEAN RATINGS FOR SIGNIFICANT INTERACTIONS
 ON THE CLASSROOM OBSERVATION SCALES (COS)

Condition	Tape Segment		
	1	2	3

(4) Percent Teacher Presentation			
High	1.78	1.69	1.53
High(+Low)	2.79	2.40	2.02

(6) Positive Affect			
High	3.28	2.88	3.08
High(+Low)	2.85	2.90	3.06

TABLE 5

Analyses of Variance of Ratings for
the Classroom Observation Record (COR)

Item	Effect (1) (degrees of freedom)			
	G (1,5)	GT (7,35)	GS (2,10)	GTS (14,70)
1 Apathetic-Alert	2.78	.30	2.86	2.14*
2 Obstructive-Responsible	.58	1.63	1.49	1.91*
3 Uncertain-Confident	.00	2.94*	.13	1.08
4 Dependent-Initiating	1.46	1.35	.77	1.63
5 Partial-Fair	.13	.54	.08	.59
6 Autocratic-Democratic	1.31	.74	1.69	.59
7 Aloof-Responsive	2.45	.45	.27	.54
8 Restricted-Understanding	.57	.88	.43	1.25
9 Harsh-Kindly	2.57	.98	2.48	.79
10 Dull-Stimulating	4.12	.40	.18	1.59
11 Stereotyped-Original	7.10*	.25	2.11	1.06
12 Apathetic - Alert	.18	.56	.01	.93
13 Unimpressive-Attractive	.01	.61	.38	.50
14 Evading-Responsible	.00	.88	1.71	.83
15 Erratic-Steady	.04	3.33**	.71	.95
16 Excitable-Poised	.10	2.66*	1.09	.97
17 Uncertain-Confident	.03	2.51*	.06	1.17
18 Disorganized-Systematic	.01	2.56*	3.09	1.18
19 Inflexible-Adaptable	.45	.23	.98	.76
20 Pessimistic-Optimistic	4.49	.99	.73	.83
21 Immature-Integrated	.23	.75	1.62	1.22
22 Narrow-Broad	.14	.44	1.01	.86
Factor (2)				
A Kindly-Understanding	2.41	.80	1.72	1.23
B Systematic-Organized	.02	2.96**	.77	1.18
C Stimulating-Imaginative	6.39*	.30	1.15	1.56
D Pupil Related Behavior	.97	1.29	1.37	1.59

* $p < .05$ ** $p < .01$

Note:

(1) G = Group: (1) Low-Inference or (2) Low- + High-Inference

T = Teacher

S = Segment

(2) See text for explanation of factor scores.

TABLE 6
 MEAN RATINGS FOR SIGNIFICANT INTERACTIONS
 ON THE CLASSROOM OBSERVATION RECORD (COR)

Condition	1	2	3	Teacher 4	5	6	7	8
(3) Uncertain-Confident								
High	5.54	5.29	4.71	5.25	6.17	5.42	5.46	6.08
High(+Low)	6.11	5.61	4.94	5.39	5.50	5.56	5.72	5.22
(15) Erratic-Steady								
High	6.13	6.04	6.04	6.33	6.21	6.17	5.83	6.42
High(+Low)	6.50	6.28	6.22	6.39	6.33	5.89	5.22	5.72
(16) Excitable-Poised								
High	6.08	5.92	6.08	6.29	6.29	6.25	5.92	6.29
High(+Low)	6.44	6.11	6.17	6.33	6.33	5.89	5.33	5.44
(17) Uncertain-Confident								
High	6.21	5.92	6.08	6.25	6.08	6.33	5.83	6.50
High(+Low)	6.61	6.33	5.83	6.17	6.17	6.17	5.06	6.28
(18) Disorganized-Systematic								
High	5.79	5.92	5.96	6.00	5.96	6.25	5.38	6.13
High (+Low)	6.44	6.22	5.94	5.89	6.11	6.00	4.89	5.56
(B) Systematic-Organized								
High	24.2	23.6	23.8	24.6	24.5	24.9	22.7	25.0
High(+Low)	25.8	24.7	24.2	24.9	24.9	23.9	21.0	22.4

TABLE 7
INTERCORRELATIONS BETWEEN FSIA CATEGORIES
AND COS SCALES*

COS Scales(1)	1	2	3	FSIA Categories(2)						
	4	5	6	7	8	9	10			
1	-06	29	-05	-25	-15	06	07	-20	17	03
2	07	28	20	18	-37	-09	02	08	10	-14
3	-12	-19	-40	-02	-06	24	16	01	44	18
4	-13	-14	-17	-22	58	-35	-04	-39	-14	-11
5	-07	-22	-08	20	05	-09	69	00	-06	11
6	-01	62	06	-28	-29	13	-21	-17	30	-04
7	01	15	08	00	-23	-08	-03	16	30	-14
8	02	-28	-11	20	26	-16	-07	08	-15	-08
9	04	11	29	06	-04	04	-06	-16	-14	-04
10	-05	35	26	-30	-05	-06	-19	-07	09	01
11	02	08	22	-10	-04	05	-29	01	06	10
12	17	24	-06	-13	-04	07	06	-23	16	02

*Decimal points omitted.

Note:

(1) See Table 3 for item descriptions.

(2) See Table 1 for category descriptions.

TABLE 8
INTERCORRELATIONS BETWEEN FSIA CATEGORIES
AND COR SCALES*

COR Scales(1)	FSIA Categories(2)									
	1	2	3	4	5	6	7	8	9	10
1	-08	18	02	-10	-10	12	11	-11	20	02
2	02	10	14	-21	-11	-05	-24	-10	07	-04
3	-02	-18	08	11	-01	-05	-16	02	09	-01
4	-05	-07	-11	-02	-13	-12	06	-22	30	-12
5	-02	03	-06	-02	00	00	-01	-20	14	12
6	03	-29	00	-01	-05	01	-12	01	12	11
7	09	20	-05	-10	-19	08	-09	-22	19	10
8	14	01	13	-17	-16	14	-19	03	06	24
9	01	-08	07	-17	-10	06	-26	06	07	23
10	-03	17	-09	-14	-10	12	22	-28	23	17
11	05	-05	-26	-07	-02	18	20	-19	16	30
12	08	12	07	-18	-11	11	01	-15	04	10
13	09	-08	19	-10	-01	18	-07	14	-11	34
14	-03	12	25	-23	-16	03	-36	-05	11	15
15	-02	19	26	-24	-11	05	-18	05	02	25
16	-03	13	27	-13	-11	08	-16	10	00	26
17	04	14		-29	-01	01	-03	-14	12	24
18	00	06	22	-17	00	-01	-14	04	-08	20
19	00	01	-03	-10	-06	06	-10	-19	27	12
20	05	02	03	-04	-09	07	-05	14	07	01
21	-06	30	18	-13	-17	07	-17	-02	16	19
22	00	-13	01	11	00	14	-05	13	-08	28

*Decimal points omitted.

(1) See Table 5 for item descriptions.

(2) See Table 1 for category descriptions.

TABLE 9
INTERCORRELATIONS BETWEEN COS SCALES
AND COR SCALES*

COR Scales	COS Scales											
	1	2	3	4	5	6	7	8	9	10	11	12
1	66	14	09	04	-03	31	21	-51	-18	12	17	40
2	41	27	-12	21	-29	33	12	-24	06	34	47	27
3	16	11	04	12	-10	14	18	-08	09	10	46	19
4	27	22	11	15	08	13	12	-15	-07	08	13	25
5	38	13	01	26	-15	31	-03	-19	09	12	29	37
6	11	04	09	06	-06	10	-01	-08	-07	06	28	15
7	40	23	-01	-04	-23	52	10	-37	-05	23	33	67
8	15	06	04	03	-24	33	02	-23	08	25	38	40
9	06	03	02	12	-25	26	03	-21	07	23	40	22
10	62	04	12	07	10	34	07	-43	-12	20	23	65
11	45	-07	27	00	13	17	03	-26	-33	02	05	55
12	46	11	02	04	-08	40	04	-32	-13	25	26	61
13	27	-01	02	02	-10	13	-01	-30	01	13	35	31
14	16	17	-11	19	-31	36	12	-15	07	47	49	22
15	12	09	-21	14	-12	35	09	-11	09	42	51	25
16	03	04	-18	09	-07	29	07	-02	06	36	40	13
17	34	12	-14	26	01	37	03	-27	-02	52	45	43
18	16	06	-29	23	-06	27	00	-15	04	54	48	25
19	25	11	02	06	-15	25	05	-16	0	22	30	43
20	35	06	03	04	-04	34	11	-27	-23	19	22	38
21	24	28	-08	08	-08	41	20	-11	07	45	34	24
22	06	01	-01	-07	09	08	10	05	-10	10	22	25

*Decimal points omitted.